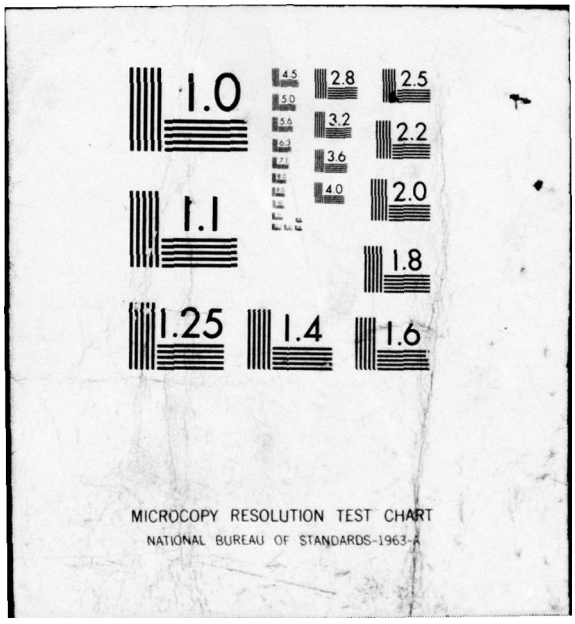
AD-A032 246 RAND CORP SANTA MONICA CALIF F/G 5/2
CLASSIFICATION OF PERSONAL INFORMATION FOR PRIVACY PROTECTION P--ETC(U)
APR 76 R TURN

UNCLASSIFIED P-5652 NL

| OF |
AD
A032246

END

DATE
FILMED
1-77


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
8.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

CLASSIFICATION OF PERSONAL INFORMATION FOR PRIVACY PROTECTION PURPOSES

Rein Turn

April 1976

23 p.

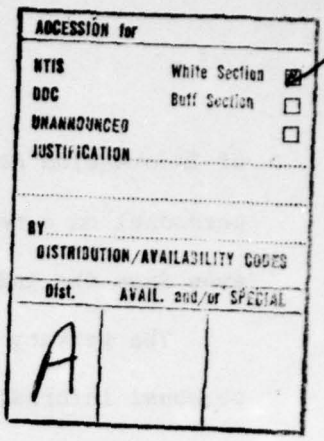
P-5652

296600

The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406



CLASSIFICATION OF PERSONAL INFORMATION FOR PRIVACY PROTECTION PURPOSES*

by Rein Turn
The Rand Corporation
Santa Monica, California

INTRODUCTION

In the United States, the federal Privacy Act of 1974[1], and similar laws enacted by other countries have established certain rights of individuals regarding personal information maintained on them by government agencies, restricted the use and dissemination of personal information, and prescribed requirements for information quality, integrity and security. In particular, the Privacy Act of 1974 states that any agency of the federal government must "maintain all records which are used by the agency in making determinations about an individual with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to assure fairness to the individual in the determination," and must "establish appropriate administrative, technical, and physical safeguards to insure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in substantial harm, embarrassment, inconveniences or unfairness to any individual on whom information is maintained." The Act and state privacy laws also note that certain personal information items are available to anyone under the provisions of the Freedom

---

*This paper was prepared for presentation at the 1976 National Computer Conference, June 7-10, 1976, New York City.

of Information Act[2], certain other items must be restricted to the agency personnel on a need-to-know basis, and still other items may be withheld even from the individual data subject himself.

The privacy protection laws recognize implicitly that not all items of personal information are equally critical in making a fair determination about an individual, that they are not equally sensitive from the point of view of their dissemination causing harm or embarrassment to an individual, and that the information quality, integrity and security requirements may vary among information items, types of records, and types of record-keeping systems. Even the same information item may be innocuous in one system of records, but very sensitive in another. For example, while a person's name is usually public information, it becomes sensitive when associated with a system of psychiatric treatment records.

Since most of the personal information record-keeping systems do not contain highly sensitive information, it is not necessary nor would it be economically practical to require absolute quality, integrity and security for all personal information in all record-keeping systems. Rather, following the approach taken in handling sensitive national defense information, a set of information sensitivity categories could be established such that, for each category, the access and dissemination restrictions would be specified and the minimal levels of required information quality, integrity and security would be defined. Thus, the technical questions of assuring quality and integrity, and providing security would be separated from the social policy questions of determining what level of integrity and protection must be provided for a particular type of information in a particular record-keeping system--several sensitivity categories are available and the

corresponding access control, integrity and security levels are provided by the system, such that any information item can be assigned to the sensitivity category most suitable under the circumstances.

This paper surveys several sensitivity classification systems that have been discussed in the literature, examines the criteria for setting up such systems, and proposes a generalized set of sensitivity categories for personal information in governmental as well as private record-keeping systems.

PROPOSED CLASSIFICATION SYSTEMS

Several suggestions for classification of personal information have been made in the literature. One of the earliest proposals by Comber[3] defines three categories based on dissemination controls that are applied:

1. Unclassified: All data maintained by a public agency not otherwise classified as restricted or confidential.
2. Restricted: Data that are not prohibited from full and free disclosure by statute (confidential), but whose unauthorized use could constitute an unwarranted invasion of personal privacy.
3. Confidential: Data that are prohibited from free and full disclosure by statutory regulation (law).

Comber suggests that among the criteria for classifying personal information as "restricted" should be whether or not the disclosure of data in question would: (1) Facilitate unwarranted identification of individuals, (2) Cause unjust economic loss or public stigma or harassment, and

(3) Result in unnecessary loss of property right. The classification decisions would be made on the basis of public policy, laws, legal interpretations, agency specifications, and personal needs of individuals. However, the decisions would be expected to vary among record-keeping systems depending on the context in which the data are embedded, the amount of information and its intrinsic nature, the sophistication of the social values of the individuals involved, and the significance of personal attributes in the sub-culture involved. Among the examples of personal information that may be classified as "restricted," Comber cites political and religious preference, marital history, family attributes, ancestry, and names of relatives.

A more detailed information classification system and sensitivity scales have been proposed by the British Computer Society[4]:

1. Public: Any personal information that is generally available in a listed form, such as various directories, biographic publications, etc.
2. Published: Information that is available but has not been collected, such as court records or hospital admission records. This category differs from "public" information in that if the individual involved does not draw attention to its existence, this information is not generally known.
3. Confidential: Information that is not generally available, although it is available and known to the individuals concerned.
4. Secret: Information that is not generally available, including the individuals concerned. Information in this category would

be collected only under statutory authority or when authorized by the individual involved.

Within these four sensitivity categories there could be a more detailed sensitivity scale, such as illustrated in Table 1. The authors of this classification system also point out that sensitivity of personal information varies with the circumstances and ideally each case should be determined according to precedents within the framework of legislation or professional codes of conduct, and that it appears not practicable to define degrees of sensitivity in any rigorous manner.

A very detailed catalog of classification of personal information items on the basis of sensitivity has been developed by Bing[5] from the point of view of Norwegian societal, legal and cultural concepts. The index contains some 400 data elements that are graded on the basis of three sensitivity levels:

1. Normal aspects (GS1): General factual information about an individual's person, family, housing, property, employment, and other information in public record-keeping systems.
2. Personal aspects (GS2): Intimate, detailed or specific information on an individual that could be used to make a social judgment about him as well as to obtain a detailed picture of his person, health, family, life style and views.
3. Disparaging and defamatory aspects (GS3): Information that could be used to form a moral or ethical picture of an

Table 1

A SCALE FOR DATA SENSITIVITY (BRITISH)

Value Scale	Examples
0	Information collected and available, such as telephone books, professional listings
1	Selected general information (e.g., titles such as Miss, Mrs. or which indicate the marital status)
2	Public utilities account inquiry systems
3	Public information in schools
4	Vehicle licensing systems
5	Financial information (e.g., bank records; medical records)
6	More sensitive financial information (e.g., company finances)
7	Commercial secure information (e.g., trade secrets)
8	Confidential police records (e.g., records used by inquiry agents)
9	Police records relating to convictions
10	Secret information (diplomatic secrets; defense secrets)

individual, as well as information on especially sensitive health conditions or handicaps, ideological views and beliefs, law enforcement information, personal idiosyncracies and habits, and evaluations of abilities.

The criteria for categorizing a data element depend, as in other classification systems that have been proposed, on considerations such as the individual and societal values, quantity of information involved, purpose of its collection and use, context and age of the information.

Sensitivity categories have also been proposed for various specialized record-keeping systems. For example, in the guidelines for record-keeping in public schools[6] the following sensitivity categories are proposed:

1. Category A: Official administrative records that constitute the minimum personal data on students necessary for the operation of the school (identification, attendance, academic work completed, level of achievement, emergency information).
2. Category B: Verified information of clear importance but not absolutely necessary (intelligence, aptitude and achievement test scores; health and family background; teacher and counselor ratings; verified reports of recurrent behavior patterns).
3. Category C: Potentially useful information, but not verified or clearly necessary beyond immediate use (legal or clinical findings, personality test results, unevaluated reports by teachers or counselors).

Specific administrative procedures are proposed for each category regarding access and dissemination, retention, and use. For example, it is recommended that unless category C items are verified and, thus, moved into category B, they should not be retained longer than one year without discussing the reasons for this with the student's parents.

Finally, the following set of sensitivity categories has been proposed for criminal justice information systems[7]:

1. Restricted: Data that requires minimum special security consistent with good security and privacy practices.
2. Confidential: Criminal justice information on individuals disseminated to criminal justice agencies, research reports derived from such information, and documentation of the information system itself.
3. Highly sensitive: Data that require maximum special security provisions and particularized privacy protection, such as criminal history information accessed by using other than personal identifying characteristics, arrest information without conviction, intelligence information, and computer programs and systems used for processing criminal justice information.

The lowest sensitivity category proposed in this system is "restricted" even though much of criminal justice information is public by statute. This is explained by pointing out that there is still a need to assure the integrity of such information.

A GENERALIZED CLASSIFICATION SYSTEM

The establishment of a standard classification system for controlling the use, dissemination and protection of personal information in record-keeping systems has the obvious benefit of clarifying for everyone concerned the level of privacy protection that can be expected and must be provided, and the consequences of not doing so. For each category would be specified the requirements for maintaining information quality, integrity and security; information handling and accountability procedures; personnel clearance criteria and procedures; information retention periods and classification criteria and procedures; and penalties for willful violations of these requirements. A framework for such a classification system is outlined below. It is discussed in more detail elsewhere[8].

A very important consideration in setting up a classification system is the number of categories that are defined. Too many categories may make the use and implementation of the system too cumbersome and costly; too few categories may result in overclassification and excessive privacy protection requirements. Important considerations here are the number of sensitivity levels of personal information and the number of different dissemination restrictions.

Sensitivity Levels

Personal information becomes sensitive when its uncontrolled dissemination may have adverse effects on the individual concerned and on his activities within his social group, or when it can reveal that the individual does

not possess values expected by his family, acquaintances, those making determinations affecting him or the society. Two situations arise: the individual wants to limit circulation of the information, or the information is kept from the individual for "his own good" or the society's good. For example, the information may include an individual's past transgressions, views or associations or, in the second case, it may include results of medical or psychiatric examinations, or information on an ongoing criminal investigation of the individual.

The adverse effects of revealing personal information on an individual to others or, as the case may be, to himself, may range from a mild annoyance to physical harm or even loss of life. Between these extremes it is possible to define many other levels of adverse effects on the individual's physical and mental health and well-being, employment, family life, reputation, social life, and values. However, in order to keep the number of sensitivity levels small a scale of six categories is proposed in Table 2. Shown are only the primary potential adverse effects of uncontrolled dissemination of information in each category; it is also possible for adverse effects to escalate into the higher categories. For example, the release of information that results in a loss of self-respect may further lead to antisocial behavior, loss of employment, and serious mental conditions.

Dissemination Categories

Another consideration in setting up a classification system involves the restrictions that are placed on dissemination of the information by statutes such as the Privacy Act and the Freedom of Information Act, and by procedures

Table 2

SENSITIVITY SCALES FOR PERSONAL INFORMATION

Sensitivity level	Potential Adverse Effects on the Individual	Examples of Information Revealed
0	No appreciable adverse effects	Widely available, common information
1	Loss of respect in social sphere, loss of friends, loss of privacy and solitude	Remarks made in private; publicly available information not widely disseminated; information on views, preferences, leisure activities
2	Loss of reputation, recognition, social acceptance, self-respect, loyalty, competence	Information on political views, anti-social behavior, evaluative statements by the individual or others
3	Loss of economic security and opportunities, employment; disruption of family life	Information on medical and psychiatric treatments; sexual deviations; extra-marital affairs; evaluative statements by or about family members; criminal history
4	Loss of civil rights, imprisonment, serious effects on mental and physical health	Self-reported information on illegal or anti-social behavior; information on medical and mental condition; psychiatric evaluations
5	Loss of life or physical safety	Information that the individual is an undercover agent for an investigative agency

adopted by the record-keeping organizations. The following possible recipients of information must be considered:

- o The individual to whom the information pertains or those formally representing his interests (guardian, physician, lawyer, accountant). There are two aspects here--knowledge by the individual that a record of information is kept on him, and access to that information.
- o Personnel of the record-keeping organization. There are two groups--those who have a specific need to use the information, and other personnel of the organization.
- o Organizations with subpoena power, such as courts, grand juries, investigative committees at various levels of government.
- o Any member of the general public who requests to see the information.

For each of the above, gaining access to the information is the principal consideration. However, for certain types of information the individual himself may need to be denied knowledge of the existence of a record on him, denied access to the content of the record, or both. For example, the Privacy Act of 1974 exempts certain testing information from access by the individual, and the existence of certain criminal investigation records may be kept secret from the individual while the investigation is in progress.

Organizations with subpoena power can demand access to any record-keeping system which they consider important for their investigation and which is not provided a privileged status by law[9]. Information that is

protected from subpoena includes the U.S. Census data and certain medical and psychiatric records. Other information granted statutory immunity from subpoenae in various states includes[10] drug abuse, alcoholism, and venereal disease records; information on victims of sex crimes, adoption proceedings, and illegitimacy records. However, personal information gathered for research purposes in social, behavioral and political sciences areas, and in education and psychology, is not provided with statutory protection against subpoenae and, as illustrated by recent cases[11,12], the researchers' promises to keep the information confidential often have no substance. Hence, every classification category should also indicate whether or not protection against supoena is provided.

Based on these considerations, dissemination control categories can range from "public" information which is accessible to anyone to "secret" information which is accessible only to authorized users of the record-keeping organization (the individual concerned neither has access to such information on him nor can he determine whether or not such a record on him exists). Table 3 depicts a classification system of seven categories that should be suitable for personal information record-keeping systems both in government and in the private sphere.

Clearly, the classification categories in Table 3 do not represent all possible combinations of access control restrictions. For example, it is conceivable that access to a particular type of information may be denied to

Table 3

CLASSIFICATION OF PERSONAL INFORMATION

Category	Individual		Access granted to users		Subject to subpoena	Access to Public	Examples of information; information sensitivity levels
	Knows	Has access	Authorized	Others			
AS: Public (by statute)	Yes	Yes	Yes	Yes	Yes	Yes	Property tax rosters; Level 0 information
A: Public	Yes	Yes	Yes	Yes	Yes	Yes	Employee directory; Level 0 information
B: Limited, Official	Yes	Yes	Yes	Yes	Yes	No	Personnel records; Level 1 information
C: Restricted	Yes	Yes	Yes	No	Yes	No	Payroll records; Level 2 information
D: Confidential (by statute)	Yes	Yes	Yes	No	No	No	Social research data; Level 3 and 4 information
E: Sensitive (by statute)	Yes	No	Yes	No	No	No	Psychiatric examination records; Level 4 and 5 information
F: Secret (by statute)	No	No	Yes	No	No	No	Organized crime investigation records; Level 5

the individual, but it may still be subject to subpoena power. In such cases the present classification system may have to be expanded.

Integrity and Security Provisions

Given a set of access control categories such as those defined in Table 3, a set of requirements for assuring integrity and security of the information in each category can be specified. However, not every category may need a separate set of specifications. In Table 3 the categories AS and A are essentially the same, and so are the categories C and D (they differ only in whether or not information in these categories is subject to subpoena power). Thus, three or four levels of integrity and security provisions may be sufficient. In any computer system there must be implemented a set of "basic" integrity, security and auditing procedures to prevent inadvertent interference of users with each other, accidental modification or destruction of information, and physical damage to the equipment[8,13-16]. Such a basic set of requirements should be sufficient for information in categories AS, A, and B (and to the sensitivity categories 0 - 2 in Table 2).

A "medium" level of integrity, security and auditing is needed for information in categories C and D (and in sensitivity levels 3 and 4) since access to this information is limited to authorized users only and this information has a greater potential for adversely affecting the individual. The security and integrity provisions should include marking of the information items and records, and hard copy, either as "Restricted" or "Confidential"; establishing users' accountability for such information; implementing more

sophisticated identification, authentication, and authorization procedures[17]; implementing audit logs; and strengthening the basic integrity assurance provisions.

A "high" level of integrity and security assurance would be provided to information in categories E and F (and for sensitivity levels 4 and 5). All files in the computer and all hard copy should be marked as "Sensitive" or "Secret" and stored securely; encryption techniques[17,19] should be used to protect information in these categories on removable storage media and in communication systems; sharing of the computer system with other computer activity should be limited, the system should be entirely dedicated to the record-keeping system in question, or (especially for sensitivity level 5) the information may have to be kept off-line or even in manual files; there should be full accountability of the users for handling information in these categories; sophisticated audit trails to trace file accesses to users and enhanced integrity control procedures such as change and error detecting codes should be implemented.

The above are only a set of general suggestions of what types of protection and integrity procedures might be used for the various information categories. In practice, the provisions adopted should reflect the specific circumstances of the record-keeping system and a thorough analysis of the security risk exposure of the information as well as a cost-benefit tradeoff. However, the methodology for risk assessment is still in the development phase at the National Bureau of Standards and elsewhere, and the cost of providing integrity and security is also known only in very rough terms[20,21].

Classification Policies and Problems

Given a sensitivity scale and a corresponding classification system it is necessary to establish a set of standard criteria and a standard policy for classifying personal information items, records that contain several information items, and entire systems of records. Certain types of personal information used by the government can be classified directly on the basis of statutes that apply to the record-keeping system or to the information categories involved. For example, all personal information collected as part of the census are automatically "confidential," while property tax records are "public by statute" and psychiatric records are "confidential" or "sensitive." For other information not covered by statutes it may be necessary to first determine its sensitivity level and then to use this for making the classification decision.

One approach to standardize the assignment of sensitivity levels is to generate a handbook where, as suggested by Bing[5], sensitivity levels are assigned to all personal information items that are known to occur in record-keeping systems (e.g., name, date of birth, amount of income, name of the employer, names of acquaintances, leisure time activities, etc.). A less detailed approach that is being considered for implementation of the Privacy Act of 1974 would assign sensitivity levels only to categories of information (e.g., identifiers, physical characteristics, employment history, evaluations, etc.) rather than individual information items and provide a list of these. In both cases there is the problem of deciding what

sensitivity level should be assigned. Sensitivity is a highly subjective and context-dependent property of personal information--what one individual may consider very sensitive may be regarded with indifference by many others, and it is likely that there is a large range of sensitivity assessments for every information item. However, it would not be practical to assign the top sensitivity level of this range to each information item. Instead, a reasonable sensitivity level must be determined through the use of surveys and expert opinion.

A traditional approach to classifying a record that contains several information items that have already been classified, or for declassifying an entire record-keeping system, is to assign to it the highest classification found among its elements. Here, too, complications may be found. For example, under some circumstances a collection of information items may be more sensitive than any one of its elements. In other record-keeping systems only very few information items may have a higher sensitivity level than the rest of the records and, thus, escalate the classification of the entire record-keeping system. The first case requires the development of an algorithm for increasing the sensitivity level of a record or record-keeping system as a function of the amount, type, sensitivity and uses of its elements. Research on this is yet to be done. The second case could be handled by using special security techniques (e.g., encryption) to reduce the sensitivity level of the information items in question, or by establishing a separate record system for these information items. Other problems that must be tackled in assigning sensitivity levels to personal information and making classification decisions include automated classification of records and

information derived from existing records, and downgrading of sensitivity levels and classifications as circumstances change.

CONCLUDING REMARKS

Laws now in force require that privacy protection be provided to personal information in record-keeping systems maintained by the federal government, and by state and local governments in three states. Other states are expected to enact similar legislation, and pending in Congress is a bill, H.R. 1984, the Comprehensive Right of Privacy Act, which would extend privacy protection also to record-keeping systems in the private sphere.

It is necessary for effective implementation of these requirements to establish a standard sensitivity classification system for personal information such that for each classification level it is known what integrity and security assurance must be provided, what information handling practices must be followed, and what penalties apply for non-compliance. Information items can then be assigned into appropriate categories on the basis of their potential to adversely affect individual data subjects, and on the basis of access and dissemination limitations that may be required by law. One such classification system and a sensitivity scale is proposed in this paper. Important questions still being researched deal with criteria and policies for determining the sensitivity and classification levels of information items, records and record-keeping systems.

REFERENCES

1. Privacy Act of 1974, Title 5, United States Code, Section 552a (Public Law 93-579), December 31, 1974.

2. Freedom of Information Act, Title 5, United States Code, Section 552, 1967.

3. Comber, E. V., "Management of Confidential Information," AFIPS Conference Proceedings, Vol. 35, 1968 FJCC, pp. 135-143.

4. Ellis, L. (Ed.), Privacy and the Computer -- Steps to Practicality, British Computer Society, London, July 1972.

5. Bing, J., "Classification of Personal Information with Respect to the Sensitivity Aspect," Databanks and Society, Universitetforlaget, Oslo, 1972, pp. 98-150.

6. Guidelines for the Collection, Maintenance and Dissemination of Pupil Records, Russell Sage Foundation, New York, 1969.

7. "*Data Sensitivity Classification*," Criminal Justice System, National Advisory Commission on Criminal Justice Standards and Goals, Washington, D.C., January 1973, pp. 128-130.

8. Turn, R., Privacy and Security in Personal Information Databank Systems, R-1044-NSF, The Rand Corporation, Santa Monica, March 1974.

9. Nejelski, P., and L. M. Lerman, "A Researcher-Subject Testimonial Privilege: What to Do Before the Subpoena Arrives," Wisconsin Law Review, Fall 1971, pp. 1085-1148.

10. "The Computerization of Government Files: What Impact on the Individual?", UCLA Law Review, September 1968, pp. 1371-1498.

11. Kershaw, D. N., and J. C. Snell, "Data Confidentiality and Privacy: Lessons from the New Jersey Negative Income Tax Experiment," Public Policy, Spring 1972, pp. 261-269.

12. Walsh, J., "Antipoverty R&D: Chicago Debacle Suggests Pitfalls Facing OEO," Science, September 19, 1969, pp. 1243-1245.

13. Guidelines for Automatic Data Processing: Physical Security and Risk Management, FIPS Pub. 31, National Bureau of Standards, Washington, D.C., June 1974.

14. Computer Security Guidelines for Implementing the Privacy Act of 1974, FIPS Pub. 41, National Bureau of Standards, Washington, D.C., 1975.

15. AFIPS System Review Manual on Security, AFIPS Press, Montvale, N.J., 1974.

16. Turn, R., and W. H. Ware, "Privacy and Security in Computer Systems," American Scientist, March-April 1975, pp. 196-203.

17. Saltzer, J. H., and M. D. Schroeder, "The Protection of Information in Computer Systems," Proceedings of the IEEE, September 1975, pp. 1278-1308.

18. Turn, R., "Privacy Transformations for Databank Systems," AFIPS Conference Proceedings, Vol. 42, 1973 NCC, pp. 589-601.

19. Feistel, H., W. A. Notz, and J. L. Smith, "Some Cryptographic Techniques for Machine-to-Machine Data Communications," Proceedings of the IEEE, November 1975, pp. 1545-1554.

20. Goldstein, R. C., The Cost of Privacy, Honeywell Information Systems, Brighton, Mass., 1975.

21. Turn, R., "Cost Implications of Privacy Protection in Databank Systems," Data Base, Spring 1975, pp. 3-9.